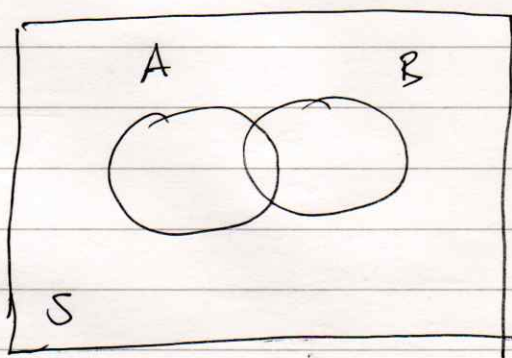


Discussion Session Week 1

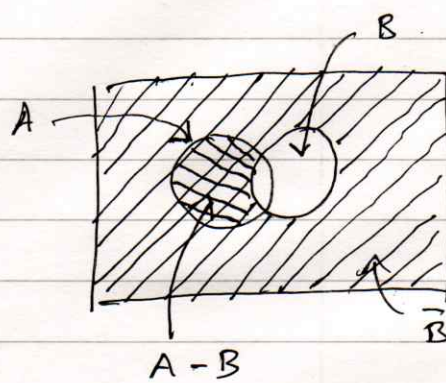
1) Example w/ Kolmogorov axioms:

$$\text{Show } P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Recall

$$A - B \equiv A \cap \bar{B}$$



$$A \cup B = (A - A \cap B) \cup B$$

↑ ↗
disjoint

$$\Rightarrow P(A \cup B) = P(A - A \cap B) + P(B) \quad (1)$$

$$\text{also } A = (A - A \cap B) \cup A \cap B \quad (\text{disjoint})$$

$$\Rightarrow P(A) = P(A - A \cap B) + P(A \cap B) \quad (2)$$

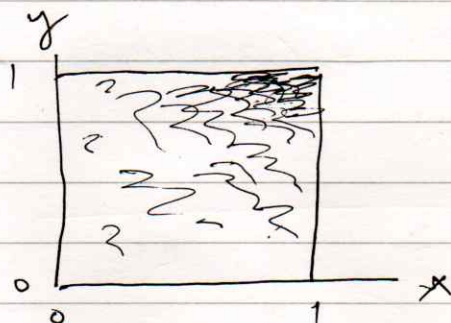
Use (1) + (2) to eliminate $P(A - A \cap B)$

$$\Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

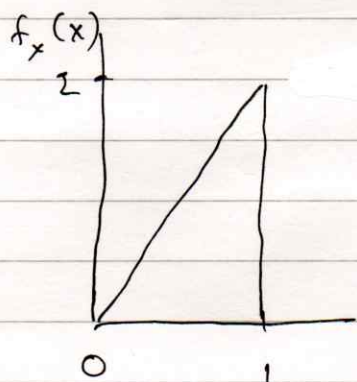
2) Example w/ joint, marginal, conditional pdfs

Consider $f(x,y) = 4xy$ $0 \leq x \leq 1$

$0 \leq y \leq 1$



• Marginal pdf $f_x(x) = \int_0^1 4xy \, dy$



$= 2x, \quad 0 \leq x \leq 1$

By symmetry,

$f_y(y) = 2y, \quad 0 \leq y \leq 1$

• Conditional pdf

$$f(x|y) = \frac{f(x,y)}{f_y(y)} = \frac{4xy}{2y}$$

$= 2x, \quad 0 \leq x \leq 1$

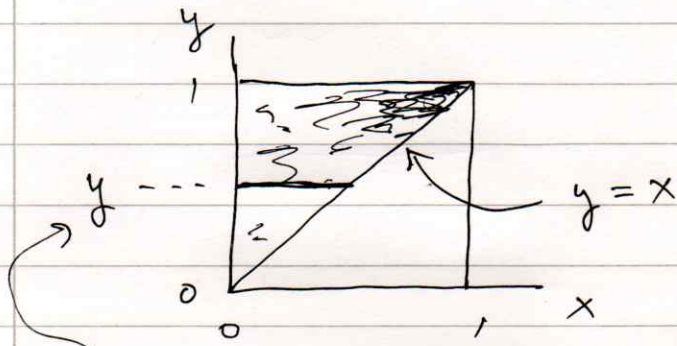
$\Rightarrow x$ & y independent ($f(x|y)$ indep. of y)

$$3) \quad f(x, y) = \begin{cases} 8xy & \\ 0 & \end{cases}$$

$$0 \leq x \leq 1$$

$$x \leq y \leq 1$$

otherwise



$$f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_0^y 8xy dx$$

$$= 4y^3, \quad 0 \leq y \leq 1$$

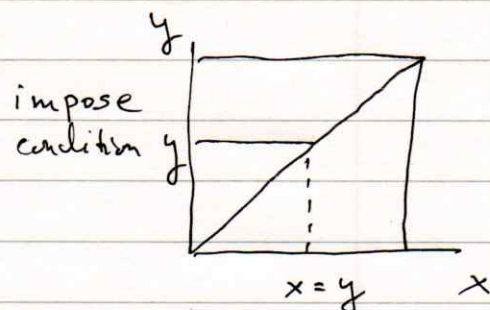
$$f_x(x) = \int_x^1 8xy dy = 4xy^2 \Big|_x^1$$

$$= 4x(1-x^2), \quad 0 \leq x \leq 1$$

3 (continued) : conditional pdfs

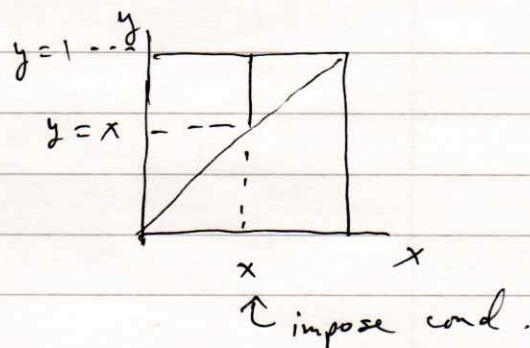
$$f(x|y) = \frac{f(x,y)}{f_y(y)} = \frac{8xy}{4y^3}$$

$$= \frac{2x}{y^2}, \quad 0 \leq x \leq y$$



$\neq f_x(x) \Rightarrow x + y$ not independent

$$f(y|x) = \frac{8xy}{4x(1-x^2)} = \frac{2y}{1-x^2}, \quad x \leq y \leq 1$$



Check Bayes' theorem

$$f(x|y) \stackrel{?}{=} \frac{f(y|x) f_x(x)}{f_y(y)}$$

$$\frac{2x}{y^2} \stackrel{?}{=} \frac{\frac{2y}{1-x^2} \cdot \cancel{4x(1-x^2)}}{\cancel{4} y^3}$$

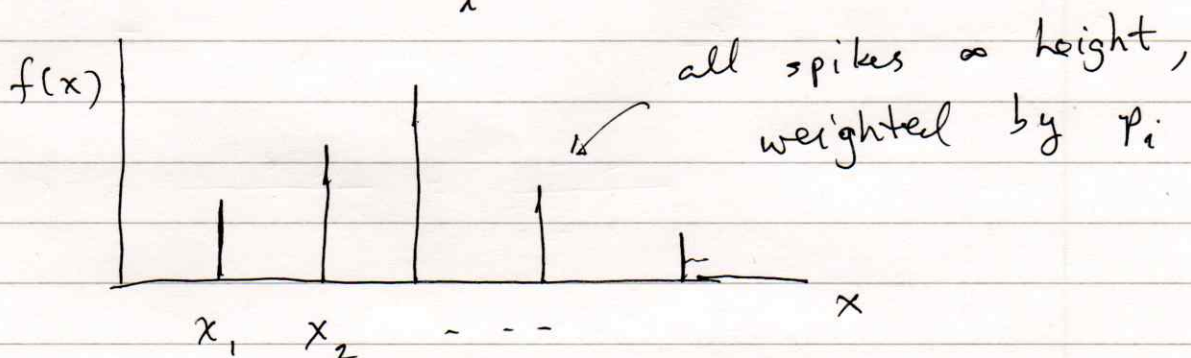
$$= \frac{2x}{y^2} \quad \checkmark$$

4) δ function \leftrightarrow pdf / cdf

Suppose discrete r.v. $x_i, i=1, 2, \dots$

$$P(x_i) = p_i \quad (\text{prob. mass func.})$$

$$\Rightarrow \text{pdf} \quad f(x) = \sum_i p_i \delta(x - x_i)$$



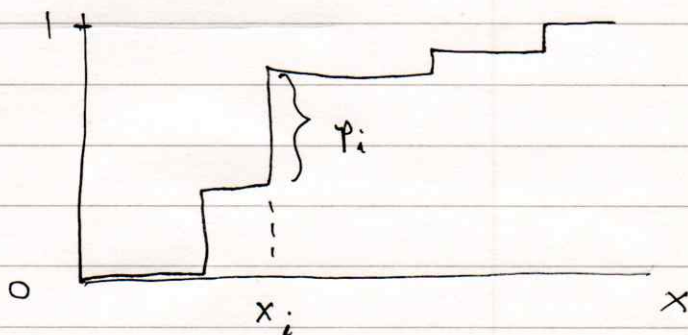
For cumulative dist., use

$$\int_a^b \delta(x - x_i) dx = \begin{cases} 1 & a < x_i < b \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow F(x) = \int_{-\infty}^x f(x') dx'$$

$$= \sum_i p_i \int_{-\infty}^x \delta(x' - x_i) dx'$$

$$= \sum_{\{i: x_i \leq x\}} p_i$$



5) Estimate # of false positives

6

$$\text{Given } \left. \begin{array}{l} P(+|D) = 0.90 \\ P(-|\bar{D}) = 0.93 \end{array} \right\} \text{ for some med test}$$

↖ has disease
↗ doesn't have disease

Out of a sample of $N = 1000$ individuals,

$$n_+ = 250 \text{ have } + \text{ test result.}$$

Estimate # of false positives in sample

$$n_+ | \bar{D} \approx P(+|\bar{D}) \cdot N_{\bar{D}}$$

↖ # w/ no disease in sample

$$P(+|\bar{D}) = 1 - P(-|\bar{D}) = 0.07$$

$$N_{\bar{D}} \approx N \pi_{\bar{D}} = N (1 - \pi_D)$$

↖ prior prob to have D

$$P(+) = P(+|D) \pi_D + P(+|\bar{D}) \pi_{\bar{D}} \approx \frac{n_+}{N}$$

↖ $(1 - \pi_D)$

$$\Rightarrow \pi_D = \frac{P(+) - P(+|\bar{D})}{P(+|D) - P(+|\bar{D})}$$

$$\Rightarrow \pi_D \approx \frac{\frac{250}{1000} - 0.07}{0.90 - 0.07} = 0.22 \Rightarrow \pi_{\bar{D}} \approx 0.78$$

$$\Rightarrow n_+ | \bar{D} = P(+|\bar{D}) N \pi_{\bar{D}} = 0.07 \times 1000 \times 0.78$$

$$\underbrace{\quad}_{\substack{\# \text{ of} \\ \text{false positives}}} = \underline{\underline{54.6}}$$